

# The Grounds of Our Freedom<sup>\*</sup>

## 1. Introduction

Harry Frankfurt's "Alternate Possibilities and Moral Responsibility" (Frankfurt 1969) was a turning point in the debates about free will and responsibility. It broke with the tradition of understanding these concepts in terms of alternative possibilities or the ability to do otherwise, and inspired and motivated a new family of views in its place: views that focus, not on alternative possibilities, but on "actual sequences" or the actual causes of behavior.<sup>1</sup> A lot has been published on Frankfurt's paper and the significance of its main argument. In this article, I focus instead on a topic that has received much less attention, but that I think is key to a correct understanding of the revisionary conception of freedom inspired by Frankfurt's argument. It concerns the connection between free will and grounding.

Why focus on grounding? A natural way to understand the question "What does it take to act freely, in the sense required for us to be morally responsible for what we do?" is as a question about the *metaphysical grounds* of our freedom (and thus, indirectly, of our moral

---

<sup>\*</sup> Thanks to audiences at Rice University, University of Cologne, Arizona State University, the 2018 Harvard/MIT graduate philosophy conference, the 5<sup>th</sup> meeting of the Latin American Association for Analytic Philosophy, the members of a graduate seminar at the University of Arizona, John Fischer, Michael McKenna, and two anonymous referees. I am very glad to be contributing to this issue in honor of Frankfurt's paper, which had such a huge influence on my work. Thanks to Geert Keil and Romy Jaster for the invitation to participate.

<sup>1</sup> Views directly inspired by Frankfurt's argument include Fischer and Ravizza 1998 (Fischer introduced the label "actual-sequence approach" in Fischer 1982: 34), McKenna 2013, and (my own view) Sartorio 2016. Frankfurt's own positive account (Frankfurt 1971) is also a view focused on actual explanations of behavior, but as I note below (see sections 2 and 5) Frankfurt's rejection of the causal theory of action suggests that he doesn't see these explanations as purely causal.

responsibility). This means that what we are after in asking this question is, roughly, an account of the facts that could *make it the case* that we act freely, or *by virtue of which* we would act freely—if we ever could.<sup>2</sup>

It may also seem natural to think that what we are after is, at least ideally, an account of the *ultimate* grounds of freedom: an account, in the most fundamental metaphysical terms, of what our freedom consists in, or would have to consist in. Some “agent-causal” views of libertarian free will are a natural expression of this idea (see, e.g., O’Connor 2000). These views suggest that our freedom is grounded in some metaphysically primitive causal powers that we possess as agents. These powers of agent-causation are irreducible to more commonplace forms of event-causation. In fact, their irreducibility is key to this conception of freedom, since it is what makes us the true originators of our choices, which on these views is essential for our having free will.

But, is an account in terms of ultimate grounds a desideratum, or at least a virtue, of a theory of freedom? This is a question that is not typically asked, but I think it’s important, and it’s the question that I want to explore in this paper. At first sight, it might seem that we will have left something substantial unexplained unless we have provided an account in terms of the metaphysically basic facts, the facts that could ultimately make us free.<sup>3</sup> Our freedom seems to be too critical to not aspire to get “to the bottom of it,” so to speak. However, I will suggest that, despite how natural this picture may appear to be, it is misguided. An account in

---

<sup>2</sup> This intuitive characterization of the grounding relation will be enough for our purposes here. The debate on the nature of grounding has become very complex in recent years, but we can safely sidestep most of it.

<sup>3</sup> Clarke (2018: 1524) seems to endorse this view.

terms of the bottom-level facts needn't be a goal of a theory of freedom. In fact, in some cases it is not even a virtue of a theory of freedom, for what we are interested in is not at all the bottom-level facts but some higher-level ones.

I will suggest that this is precisely what happens with the Frankfurt-inspired views, the views in terms of actual sequences or actual causes.<sup>4</sup> I will argue that we shouldn't expect these views to shed any light on the ultimate grounds of freedom. For, as we will see, doing so misrepresents their nature and unfairly limits their potential. Although this may be true as well of other theories of freedom, the Frankfurt-inspired views are particularly noteworthy in this respect because of the kind of theories that they are (causal theories *par excellence*, as I will suggest we call them in section 2), which can naturally result in confusions about their ambitions. Hence my special focus on these views.

In particular, as we will see, it is common to characterize these new views as aspiring to be non-modal or non-counterfactual accounts of freedom. I will show that this characterization is mistaken, and in potentially significant ways. For counterfactual facts *can* consistently play a role in these views, insofar as those facts can act as grounds for the relevant facts about actual causes. The reason this isn't obvious at first sight is, again, that the main target of a theory of this kind is not the lower-level grounding facts but the higher-level ones (the facts about actual causes themselves), which can obscure the truth about these matters. Thus, the paper is an attempt to clear up these confusions, and to draw other important implications for debates about free agency.

---

<sup>4</sup> Note that these are not agent causes, as in those libertarian views referred to above, but ordinary event-causes.

The structure of the paper is as follows. In section 2, I introduce the labels “causal theories” and “causal theories *par excellence*,” and I characterize the Frankfurt-inspired views in terms of those labels. I also explain the relation, as I see it, between causal theories of freedom and causal theories of action. In section 3, I discuss grounding—and, in particular, the potential role played in causal theories (of action and freedom) by the grounds of causal facts. This section contains the main argument that the Frankfurt-inspired views of freedom target the higher-level facts, not the bottom-level grounding facts. In sections 4 and 5, I draw some important implications of the previous discussion. Section 4 concerns the contrast between the Frankfurt-inspired views and the traditional views of freedom, and section 5, another debate fueled by a paper by Frankfurt: the debate between Frankfurt and Mele on the phenomenon of “passive agency.” I end with some concluding remarks in section 6.

## **2. Causal theories and causal theories *par excellence***

Let me start with a more thorough characterization of the views of freedom that will be our focus here. These views start from the idea that freedom (again, the kind of freedom relevant to moral responsibility) is just a function of the actual explanation of our behavior, or of why we in fact do what we do. As noted above, this idea was forcefully suggested by Frankfurt in his 1969 paper. There Frankfurt argued that responsibility (more precisely, the kind of freedom required for us to be responsible) is not a matter of our being able to do otherwise, but it’s just a function of the factors that actually explain our behavior.

Frankfurt illustrated his argument with scenarios (now commonly called “Frankfurt-style” cases) involving agents who, despite lacking the ability to do otherwise, still seem to act

freely and be morally responsible for what they do.<sup>5</sup> The most popular cases of this kind involve a nefarious and resourceful neuroscientist who is secretly monitoring the deliberation of an agent, Jones, and who could have intervened by manipulating his thought processes so as to make him decide to do what he wants him to do. As it turns out, however, the neuroscientist never has to intervene because Jones acts on his own, on the basis of his own reasons, and in exactly the way the neuroscientist wanted.

Frankfurt's reasoning rests on the following insight: the presence of the neuroscientist is completely irrelevant to Jones's freedom because it doesn't affect the actual explanation of his behavior in any way. Imagine that Jones were to find out later about the neuroscientist's presence and tried to excuse his behavior by drawing attention to that fact. Intuitively, this would seem inappropriate, for the neuroscientist played absolutely no role in accounting for his behavior. Thus, Frankfurt argued, this motivates the idea that the only factors that are relevant to Jones's freedom are those that have to do with the actual explanation of his behavior, or with why he did what he did.

Note that the most natural way to put this idea is in terms of causes: the only factors that are relevant to an agent's freedom are those that concern the *actual causes* of the agent's behavior. Interestingly, this is not Frankfurt's own preferred way of putting it, for reasons to which we will return later, and that have to do with his rejection of the causalist view of agency (more in section 5 below). But since the causalist view of agency is by far the most widely accepted view these days (and for good reasons, I think), and since the most natural

---

<sup>5</sup> I discuss the literature on Frankfurt-style cases in Sartorio 2017a.

interpretation of Frankfurt's insight is in terms of actual causes, I suggest that we adopt this interpretation from now on and see what follows from it.

One interesting thing that follows is that views of freedom based on this insight turn out to be views that can be characterized as *causal theories of freedom*. In general, causal theories are views that focus on actual causes or actual causal connections.<sup>6</sup> Think about other examples of (philosophical) causal theories. There are plenty to choose from! A causal theory of knowledge (Goldman 1967) states that in order for a belief in *p* to count as knowledge there must be an actual causal connection (of the right kind) between *p* and the belief that *p*. Similarly, a causal theory of reference (Kripke 1980) states that in order for a name to refer to an object there must be an actual causal chain (of the right kind) linking the name and the object. And a causal theory of action (Davidson 1963) states that in order for a behavior to constitute an action the behavior must have a certain kind of actual causal history, one that includes the relevant mental items. All of these views introduce a key causal condition (a condition that needs to be satisfied for a belief to count as knowledge, or for a name to refer to an object, or for a behavior to count as an action). This condition concerns actual causal histories.

Some causal theories are what I will call "causal theories *par excellence*." These views don't just include a key causal condition; in addition, they also claim that actual causal histories are *all that matters*. Both causal theories of action and causal theories of freedom are of this kind (I won't try to adjudicate the other examples of causal theories discussed above). A causal

---

<sup>6</sup> "Causalist" is another term that is sometimes used to refer to these kinds of views, especially in the case of action. Here I'll use the two labels interchangeably.

theory of action is a theory that states that all it takes for a behavior to be an action is to be caused, in the right way, by certain mental antecedents: belief/desire pairs, intentions, etc. (or perhaps by their physical realizers, if mental causation is a problem). In other words, actual causal histories are all that matters for a causal theory of action. In turn, the original motivation for causal theories of freedom suggests that we should also understand them as causal theories *par excellence*. For, again, the initial motivation is the idea that, if a factor plays no role in the actual causal explanation of the agent's behavior, it is irrelevant to the agent's freedom. In other words, again, actual causal histories are all that matters.

Not only are causal theories of action and causal theories of freedom both causal theories *par excellence*; in fact, there is a natural way to understand one as an extension of the other. A causal view of action postulates some basic requirements that the causal history of a behavior needs to satisfy in order for it to be an action. In turn, a causal view of freedom can be seen as positing some additional requirements that the causal history needs to satisfy in order for the action to be done freely. For example, a causal theory of freedom will have to explain why it is that a behavior done out of compulsion (such as compulsive hand-washing) is not a free action, although it may be an action. The explanation will have something to do with the actual history of the compulsive action and how it differs from the causal history of a free action. In so doing, the theory will appeal to requirements that go beyond the requirements for a behavior to constitute an action.<sup>7</sup>

---

<sup>7</sup> In Sartorio 2016 (chapter 4) I argued for a way of doing this in terms of causes that reflect the agent's sensitivity to reasons when the agent acts freely.

### 3. Grounding

Now let's think about how a causal theory of freedom can express its central claim in terms of grounding.<sup>8</sup> Again, as a causal theory *par excellence*, a theory of this kind states that freedom is just a function of actual causal histories: all that matters are actual causes. On a first pass, then, the grounding claim seems to be the following:

**(G)** Facts about freedom are *exclusively grounded in* facts about actual causes.

But how exactly should we understand G and its inherent “exclusiveness” claim? Take the example of Jones and the neuroscientist. Imagine that we want to say that a certain freedom fact, such as the fact that Jones acted freely at *t*, is exclusively grounded in facts about the actual causes of Jones's act. Imagine, for example, that we want to say that it is exclusively grounded in the fact that Jones went through a reasons-responsive process of deliberation that causally resulted in his behavior, in the right kind of way (this is just an example of what a causal theory might say, and it will do for our purposes here). Now consider the following two facts:

**Freedom fact:** Jones acted freely at *t*.

**Causal fact:** Jones's reasons-responsive process of deliberation caused his behavior at *t*, in the right kind of way.

---

<sup>8</sup> Part of what follows is an expansion of some ideas developed in Sartorio 2016, chapter 1. For a discussion of some related points, see Sartorio 2018a (especially my reply to Moya 2018 and Whittle 2018) and Sartorio 2018b (especially my reply to Clarke 2018 and Pereboom 2018).



A natural way to read G, and the exclusiveness claim inherent in G, is simply as the claim that the causal fact grounds the freedom fact, and that the causal fact is the *only* fact that grounds the freedom fact. In other words, *no other* facts besides the causal fact ground the freedom fact.

But this is not how G should be understood. Think about other causal views and how we understand the grounding claims that they make. Take our other main example of a causal theory *par excellence*: a causal theory of action. Imagine that Mary raised her arm at t, and that we want to say that Mary's raising her arm is an action, and not merely something that happens to her (a mere "arm rising") because it has the right kind of causal history—say, it is caused (in the right way) by Mary's intention to raise her arm. Consider the following facts:

**Agency fact (A):** Mary's behavior at t is an action.

**Causal fact (C):** Mary's intention caused (in the right way) her behavior at t.

As a causal theory *par excellence*, a causal theory of action claims that facts about agency are, again, "exclusively" grounded in facts concerning the actual causal history of behavior. But, does this mean that a fact like A is grounded in a fact like C and in no other facts?

Clearly not. Note that C is not any old fact; it is a *causal* fact. And a causal theory of action is (in and of itself) neutral on the issue of the grounds of causation. It is obviously not the job of a causal theory of action to shed light on the nature of causation! So, a causal theory of action will leave it open whether C is itself grounded in other facts, and what those other facts

might be. But note that, if C were grounded in other facts, then it would arguably follow that other facts besides C itself ground A, namely, the grounds of C.<sup>9</sup>

In sum, C's being a causal fact, and so a fact that could at least in principle be grounded in other facts, opens the door to additional facts besides C itself that could be playing a role in grounding A. Thus, the claim that a causal theory of action makes, according to which facts about agency are "exclusively" grounded in facts about the actual causal history (i.e. the claim that makes it a causal theory *par excellence*), should be interpreted in a way that is consistent with this. Arguably, it should be interpreted in a way that entails that facts like A are only grounded in facts like C *and*, at least potentially, the grounds of C.

In retrospect, this is the only interpretation that makes sense. If one claims that something, say, a very clean lake, is exclusively made out of water, one is not thereby denying that the lake is made out of the components of water itself (hydrogen and oxygen, and electrons, protons, quarks, and so on). Similarly, in our case: the claim that a fact like A is exclusively grounded in a fact like C should be interpreted in a way that is consistent with the grounds of C also grounding A.

The same goes for a causal theory of freedom, then. Claim G above should be understood in a way that makes room for the causal facts themselves to be grounded in other facts, and thus for other facts to play a role in grounding the freedom facts. Arguably, it should

---

<sup>9</sup> Unless this is one of those rare instances where (some metaphysicians would say) the transitivity of grounding fails; see, e.g., Rodriguez-Pereyra 2015. I seriously doubt this. But, at any rate, a causal theory of action is, in and of itself, neutral on this issue as well; it is not the job of such a theory to adjudicate this issue either. So, again, the theory would have to leave it open whether there are other facts that ground the action facts; for all the theory says, there *might* be some such facts.

be understood as the claim that facts about freedom are only grounded in actual-causes facts *and*, at least potentially, the grounds of those actual-causes facts.

We may now return to our original question: Is an account in terms of ultimate grounds a desideratum or at least a virtue of a theory of freedom? It should be clear now that this is not at all a desideratum or even a virtue of, in particular, a *causal* theory of freedom. In giving a causal theory of freedom, we are not aiming to provide an account in terms of the bottom-level facts. In fact, we are typically trying to remain neutral (or as neutral as possible) on the bottom-level facts. For, again, it is not part of the job of a causal theory of freedom to elucidate the concept of causation, in the same way it is not part of the job of a causal theory of knowledge or reference or action to elucidate it. In general, in formulating a causal theory of a certain concept, we are just interested in the higher-level facts—the causal facts themselves—and not in the grounds of those causal facts. Plainly, in formulating such theories we can make use of the concept of causation without having to rely on any particular metaphysical account of it.

Still, at the same time that we recognize that our focus is the higher-level facts, we also shouldn't lose sight of the possibility that these facts *might* be grounded in other facts. Otherwise, as we have seen, this results in a misunderstanding of a causal theory *par excellence* and what its "exclusiveness" claim really amounts to. So, although our focus is not the bottom-level facts, we should remain aware of the possibility that some lower-level facts might exist and do some grounding work of their own.

This reasoning extends to other accounts of freedom beyond causal theories. Although other theories are not my main focus here, I think it's clear that the same applies to views that make use of other metaphysical concepts without relying on their being irreducible. I have in

mind, for example, theories that rely on potentially reducible concepts such as certain kinds of abilities, dispositions, etc. But the case of causal theories of freedom is particularly interesting given the exclusiveness claim that they are committed to, in light of their being causal theories *par excellence*. For, as we have seen, this aspect of the views can naturally give rise to misunderstandings about their aspirations.

In the rest of the paper I'll discuss some implications of the previous discussion.

#### **4. First implication: The debate between alternative-possibilities views and causal views**

One implication concerns how we should think about some free will debates. Let me explain.

As noted above, causal theories of freedom originated in opposition to the classical model that was Frankfurt's main target: while the classical model required alternative possibilities of action, the new theories proposed that all that matters is the relevant facts about actual sequences or actual causes. Now, over the years this has resulted in some confusion over what exactly is being debated between the two kinds of views. In particular, it is quite common to see the new theories characterized as "non-modal" attempts to account for our freedom, or as views that have turned the focus away from counterfactual facts (facts concerning possibilities, abilities, or the like) and onto facts about actual causal sequences exclusively.<sup>10</sup>

---

<sup>10</sup> For example, in discussing the Fischer and Ravizza view, Levy (2008: 223) characterizes their view as a view according to which "moral responsibility depends only upon actual sequence properties of agents, not on what might happen in various counterfactual scenarios." Similarly, Clarke (2018: 1517) characterizes Sartorio's view as "a reasons-sensitivity theory, one on which the relevant sensitivity is construed not modally or counterfactually but in terms of actual causation." Franklin (2016), Tognazzini (2016), Kearns (2017), and Bernstein (2018) also seem to think that causal views are views that aspire to avoid any commitment to

But, as we have seen, this is a misconception. For causal theories should *not* be interpreted as claiming that the only grounding facts are actual causal facts; instead, they should be interpreted as claiming that the only grounding facts are actual causal facts and, at least potentially, the grounds of those actual causal facts. Consequently, a causal theory of freedom can consistently accept the significance of some counterfactual facts, insofar as these facts can contribute to the grounding of the actual causal facts.<sup>11</sup> In fact, as it turns out, many contemporary views of causation are *counterfactual* in nature—they claim that causal facts are grounded (at least partly) in counterfactual facts.<sup>12</sup> As a result, if any of these views of causation were true, then some counterfactual facts would ultimately play a role in grounding freedom, according to causal views of freedom.

What, then, is the right way of understanding the contrast between the classical view of freedom and the views based on actual causes? I think that we should understand it, not in terms of the relevance or irrelevance of counterfactual facts generally speaking, but in terms of the *specific* kinds of facts at issue.

---

counterfactual or modal notions. For a related discussion (on what kinds of abilities/counterfactual possibilities the classical view and the causal view might appeal to), see Franklin 2015 and Cyr 2017.

<sup>11</sup> The claim that these views of freedom are consistent with the relevance of some counterfactual facts is not new (notably, the Fischer and Ravizza view relies on the truth of this claim too). What's new is the *justification* offered for this claim. The justification I'm offering is something that we can only see clearly upon reflecting on the role played by grounding in our conceptions of freedom. Thanks to John Fischer for prompting me to make this clarification.

<sup>12</sup> The classical example is Lewis 1986. For a discussion of more recent views of this kind, see, e.g., the papers in Collins, Hall, and Paul 2004. There is now close to a complete consensus that causation doesn't reduce to simple counterfactual dependence. Accordingly, contemporary counterfactual views cash out causation in terms of counterfactuals that are much more complex in kind—for examples, counterfactuals that hold fixed certain facts about the actual circumstances.

Here is why. Arguably, what characterizes the conception of freedom in terms of alternative possibilities is not *just* a commitment to the relevance of some counterfactual facts, but a commitment to the relevance of a certain kind of counterfactual fact. For the mere existence of counterfactual scenarios where agents do otherwise is not enough, by itself, for those agents to be able to do otherwise in the required sense. What's missing, intuitively, is something more *robust*, such as the agents' having access to those counterfactual possibilities, or its being within their power to actualize them. In the free will literature, this "robustness" of the alternative possibilities has been discussed mostly in connection with the debate about Frankfurt-style cases and whether they successfully undermine the classical model (see, e.g., Pereboom 2014, chapter 1). But what this suggests is that the robustness of alternative possibilities is also needed, more fundamentally, to understand the very terms of the debate itself: it is needed to understand *what exactly is being debated* between the advocates of the alternative-possibilities model and their critics.

My proposal, then, is that the central question that is being debated is not whether freedom is grounded in counterfactual facts (since both views are compatible with this), but whether freedom is grounded in a certain class of counterfactual facts that represent robust alternative possibilities of action. For, again, the question is not simply whether, in order for us to be free, there should be counterfactual scenarios where we do otherwise. It is, rather, whether, in order for us to be free, there should be counterfactual scenarios where we do otherwise that are accessible to us in the relevant sense, or that it is within our power to actualize.

I have argued that the right way of understanding the grounding claim about freedom made by a causal theory *par excellence* has important implications about how to understand certain fundamental debates about free will. Now I will argue that it also has important consequences for how we should adjudicate some of those debates. In the next section, I illustrate this point with an analysis of a particular example (which I think is representative of the broader set of issues discussed in this paper): the debate between Frankfurt and Mele on the phenomenon of “passive agency” and the implications for causal theories of action and free action.

## **5. Second implication: The debate about passive agency**

As anticipated earlier, Frankfurt would object to a causal interpretation of the idea that freedom is just a function of the actual explanation of behavior, given his rejection of causalism as a theory of agency. Clearly, if the causal account of action fails, then so does a causal account of free action (since, as explained above, causal accounts of free action can be regarded as extensions of the causal account of action). And Frankfurt thought that there are good reasons to reject the causalist view of agency. He argued that a behavior is an action just in case it is “under the agent’s guidance,” where this is a concept that cannot be adequately expressed in causalist terms (Frankfurt 1978).

An important part of Frankfurt’s argument against the causalist view of agency involves an example of the following kind:

**Downhill Coasting:** An agent is driving his car when he gets to the top of a hill. The car starts coasting downhill as a result of gravitational forces alone. Since the driver is completely satisfied with the direction and speed of the car, he doesn't intervene at all during that time (we may even imagine that he takes his hands off the wheel and his feet off the pedals); he simply stands ready to intervene if necessary, while closely monitoring the course of the car.

Frankfurt argued that in this case the movement of the car remains under the driver's guidance without the driver or the driver's intentions being part of the actual causal chain. The agent *would have* causally affected the course of the car if any adjustments had been needed, but in fact he doesn't exert any such causal influence, since no adjustments were needed. In other words, the type of control that the agent enjoys in this case seems to involve purely counterfactual causal connections and no actual causal connections.<sup>13</sup> Thus, Downhill Coasting appears to be a problem for causalism about agency because it suggests that agents can remain in control and exercise their agency in the absence of any actual causal connections of the kind the causalist view would require.

Mele (1997, section 2.D) replied to Frankfurt on behalf of causalism. He argued that Downhill Coasting (a case of "passive" agency, as he called this interesting form of agency) is not a problem for causalism because there is in fact some intention of the agent causally sustaining the behavior of the car throughout that time, when he is allowing the car to go

---

<sup>13</sup> Frankfurt then applied the same argument to the relation between an agent and his bodily movements. In what follows I'll work with the Downhill Coasting case instead, since this is the case discussed by Mele.



downhill in the precise way it does. This intention is among the causes of the car's behavior: the car is moving in a certain way partly *because* the agent wants it to. I take it that what Mele has in mind is that, although there are other causes of the car's behavior (such as, notably, gravity), the driver's intention to go downhill (in a certain direction, at a certain speed, etc.) is one of those causes. Moreover, it is in virtue of this fact that the agent remains in control of the car: if it were not for that causal connection between his intention and the car's movements, the driver would not be in control of the car. As a result, Downhill Coasting (and the phenomenon of passive agency in general) is not a counterexample to the causalist view.

Mele anticipates how Frankfurt might respond to this suggestion and offers a counter-reply. He notes that Frankfurt will likely insist that the control that the driver has of the car in that case doesn't stem from any *actual* causal connections but from purely *counterfactual* causal connections—that is, from the fact that he *would* have been able to intervene more or less effectively if that had turned out to be necessary. And Mele's counter-reply to this is that this cannot be right because such counterfactuals are, in fact, irrelevant. Applying the same strategy that Frankfurt himself popularized in the free will debate, the strategy of appealing to the presence of idle counterfactual interveners, Mele imagined a version of the case where a mind-reading demon who had been closely monitoring the driver's thoughts and movements wouldn't have let the driver make any changes to the course of the car (if he had tried to make changes, the demon would have paralyzed him completely until the car ran its course down the hill). Mele claimed that the driver remains in control of the car even in a case like this, despite the fact that he lacks the counterfactual type of control imagined by Frankfurt—that is, even if it's not true that, had he tried to make certain adjustments to the course of the car, he would

have successfully made them. Thus, Mele argued, the only kind of control that matters is actual causal control, and causalism is, once again, vindicated.

Several things go wrong in this exchange, I think, and our discussion in the preceding sections can help explain why and shed some useful light on this debate—ultimately by defending causalism, but in what I think is a more successful way of doing that. On the one hand, as Mele’s reply helps to bring out, Frankfurt seems to be inferring from the fact that the driver has a kind of counterfactual control that the car’s behavior is under his guidance in a way that doesn’t reflect on its actual causes, and this inference seems unjustified. On the other hand, Mele seems to want to drive a wedge between causal and counterfactual connections that is not needed to give an adequate defense of causalism. Plus, driving that wedge might be a bad idea for other reasons.

Let me explain. As we have seen, although a causal theory of (free) action relies on actual causal connections, it leaves open what those causal connections may be grounded in, if anything. In particular, for all a theory of that kind says, causal facts *may* be grounded in counterfactual facts, at least partly. So, even though Frankfurt might be right that the agent’s control in Downhill Coasting has something important to do with the truth of certain counterfactuals, Mele could still be right that Downhill Coasting isn’t a problem for causalism, for those counterfactual facts could be supporting the relevant causal facts. In that case, the agent’s intention would still be causally connected with the car’s behavior in a way that allows for the agent to be in control, in accordance with the causalist view.

To illustrate this point, let’s distinguish the following three facts:

**Agency fact:** The driver's coasting downhill at T is an action.

**Causal fact:** The driver's intention caused (in the right way) the car's movements at T.

**Counterfactual fact:** There is the right kind of counterfactual connection between the driver's intention and the car's movements at T.

What I am suggesting is that the following grounding structure is compatible with causalism about agency: the agency fact is grounded in the causal fact, which in turn is grounded in the counterfactual fact; as a result, the agency fact is also grounded in the counterfactual fact. (Of course, the grounding chain needn't stop there; note, in particular, that, if actualism about modality were true, then the relevant counterfactual fact would itself be grounded in actual facts.)

Now, Mele seems to think that these kinds of counterfactual facts don't play any such grounding role. But his reasons for thinking this don't seem to me convincing. Again, he suggests that those counterfactuals are irrelevant because when we take away the counterfactual connection that Frankfurt was imagining existed between the agent's intentions and the car's movements we don't thereby take away the control of the agent. In particular, he thinks that the driver would remain in control of the car even if it turned out that he wouldn't have been able to make any such adjustments if he had tried (because the mind-reading demon would have paralyzed him).

However, it seems to me that the presence of the demon *would*, in fact, undermine the driver's control of the car: it seems to me that the agent would have no control of the car, in the relevant sense, in a case of that kind. If this isn't clear at first sight, it could be because even

in this case it's undeniable that the driver bears some *indirect* causal connection to the movements of the car. After all, he started the car in the first place, and the presence of an idle demon in the background cannot change that. But what the causalist needs in order for the driver to stay in control of the car at that later time, once the car starts going downhill, is a more direct kind of causal connection between the intention to go downhill and the car's going downhill. And it's much less clear that such a causal connection exists, if the driver was unable to make any corrections to the course of the car.

To test this hypothesis, let's remove from our example any indirect causal connections that might have existed between the driver and the current movement of the car. Consider, for example, the following new version of the case:

**Teleported driver:** As the car starts going downhill, a random fluctuation in the fabric of space-time results in the driver of the car being teleported to a distant location and his being replaced with a new, also teleported, driver who is now occupying the driver's seat (hands off the wheel and feet off the pedals).<sup>14</sup>

Imagine that everything else stays the same, so the new driver feels no need to make adjustments to the course of the car as it's coasting downhill, because the car is doing exactly what he wants it to do. However, he wouldn't have been able to make any changes even if he had tried, because (unbeknownst to him) the mind-reading demon wouldn't have let him. I

---

<sup>14</sup> Although the teleportation makes for a "cleaner" case, if it helps the reader one could also imagine a more ordinary case where two people in the car just switched places and there is a new driver in front of the wheel now. That case elicits similar intuitions in me.

think it's clear that our new teleported driver has no control whatsoever of the car. Although it will surely appear to him as if he is in control, he is in fact not causally linked in any way to how the car is moving. As a result, he is *not* performing an action of coasting downhill.

The teleported driver finds himself in a situation relevantly like the situation of the agent in this case:

**Sharks:** John is walking along a beach when he sees a child drowning in the sea. John believes that he could easily save the child, but he decides not to help him, and continues to stroll along the beach. The child drowns. As it turns out, however, John couldn't have saved the child because some hungry sharks would have eaten him, had he jumped in. (This is a case discussed in Fischer and Ravizza (1998: 125)).

As Fischer and Ravizza note (and pretty much everybody tends to agree), John isn't responsible for the child's death in this case, because, despite what he had reason to believe, the fate of the child was not at all in his control. For similar reasons, I am arguing, the teleported driver isn't at all in control of the fate of the car. Although he has good reason to believe that he is in control, he is just wrong about this, in the circumstances.

This suggests that, *at least in cases of this kind*, counterfactual connections, even the simple or more straightforward counterfactual connections that Frankfurt originally had in mind, may be playing more of a role than Mele thinks in grounding action, and thus free action. For, in those cases, the counterfactuals seem to be reliably tracking whether the agent is in control of the relevant outcome. Without getting into all of the details here, the reason for this

has probably something to do with the fact that the phenomenon of passive agency is an unusual case of action in that it involves, well, *in*action. (Hence the label “passive”: if the agent acts, it is by remaining passive, in some important sense, or by omitting to act...) And causation involving omissions is quite generally thought to be more closely connected to counterfactual dependence than other forms of causation.<sup>15</sup>

On some views (notably, Hall 2004), causation involving omissions is even its own variety of causation, and one that is simply identified with counterfactual dependence. But we don’t have to embrace this extreme view to realize that counterfactuals seem to be playing more of a role when we’re dealing with causation involving omissions than in other cases. If you couldn’t have saved the life of a child who drowned in the sea because you would have died trying, it seems clear that in no way did you cause his death by not attempting a rescue. This is arguably due to the failure of counterfactual dependence: you wouldn’t have saved him, even if you had tried. In contrast, an assassin clearly causes the death of his victim when he shoots her at point blank, even if the victim would still have died in a very similar way in the hands of a backup assassin if he hadn’t shot her first. (In fact, the Frankfurt-style case of Jones discussed in section 2, which we used to illustrate the motivation for the causalist view of freedom, is of this kind: although the presence of the neuroscientist results in the effect being overdetermined

---

<sup>15</sup> There is a complication here in that some views don’t allow for any causation by omission. I am setting those views aside because they are not the kinds of views that a causalist about agency would probably endorse anyway, since they lack the basic causal elements to account for the type of agency involving omission. (I discuss this and related issues in Sartorio 2009.) There might be ways of extending those views by utilizing quasi-causal concepts, or by appeal to the concept of causal explanation. On the use of surrogate concepts such as these to accommodate omissions within an actual-sequence theory of freedom, see Sartorio 2016: chapter 2.

and thus in a failure of counterfactual dependence, this isn't enough to break the relevant causal connection involving Jones, which is why Jones still remains in control and can be responsible for what he does.)<sup>16</sup>

In sum, counterfactual connections seem to be playing an important role in cases of passive agency. But, again, what's important to bear in mind is that this isn't a problem for causal views of action or free action. Instead, it's a reminder of the kind of project we embarked on when we first thought of grounding free action in actual causal histories. As I have argued, what we are after in formulating a theory of this kind is, not an account in terms of the lower-level grounding facts, but an account in terms of the higher-level facts, the causal facts themselves. At the same time, however, it's an account that *leaves room* for some lower-level facts to act as grounds for the higher-level facts. And these lower-level facts could, in principle, be counterfactual facts of certain kinds (indeed, on counterfactual views of causation or of causation by omission, they will be).

## 6. Conclusions

To conclude, we have examined some aspects of the project of grounding freedom. I have argued that a theory of freedom shouldn't necessarily aspire to be an account of the ultimate grounds, since sometimes what we are interested in is not the bottom-level metaphysical facts but some higher-level facts. I have suggested that this is the case, in particular, with the views that were inspired by the publication of Frankfurt's 1969 paper: views that understand freedom in terms of actual sequences or actual causes.

---

<sup>16</sup> For more on causal differences of this kind, see Sartorio 2017b.

I have also argued that reflecting on these issues about grounding can be fruitful in two important ways. First, it can help us advance our understanding of certain debates about free agency by shedding clarity on what is being debated and by avoiding some common confusions. And, second, it can help us make progress in adjudicating some of those debates, by allowing us to identify the moves that are legitimately available to each side.

## References

- Bernstein, S. (2018) Review of *Causation and Free Will*, by C. Sartorio. *Philosophical Review* 127, 4: 550-4.
- Clarke, R. (2018) "Free Will, Causation, and Absence." *Philosophical Studies* 175, 6: 1517-24.
- Collins, J., N. Hall and L. A. Paul (eds.) (2004) *Causation and Counterfactuals* (Cambridge, MA: MIT Press).
- Cyr, T. (2017) "Semicompatibilism: No Ability to Do Otherwise Required." *Philosophical Explorations* 20, 3: 308-21.
- Davidson, D. (1963) "Actions, Reasons, and Causes." *Journal of Philosophy* 60: 685-700.
- Fischer, J. "Responsibility and Control." *Journal of Philosophy* 79, 1: 24-40.
- Fischer, J. and M. Ravizza (1998) *Responsibility and Control* (Cambridge: Cambridge University Press).
- Franklin, C. (2015) "Everyone Thinks that an Ability to Do Otherwise is Necessary for Free Will and Moral Responsibility." *Philosophical Studies* 172, 8: 2091-2107.



- Franklin, C. (2016) Review of *Causation and Free Will*, by C. Sartorio. *Notre Dame Philosophical Reviews*.
- Frankfurt, H. (1969) "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy* 66: 829-39.
- Frankfurt, H. (1971) "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68: 5-20.
- Frankfurt, H. (1978) "The Problem of Action." *American Philosophical Quarterly* 15: 157-62.
- Goldman, A. (1967) "A Causal Theory of Knowing." *Journal of Philosophy* 64, 12: 357-72.
- Hall, N. (2004) "Two Concepts of Causation." In Collins, Hall, and Paul (2004), pp. 225-76.
- Kearns, S. (2017) Review of *Causation and Free Will*, by C. Sartorio. *Ethics* 127, 3: 802-6.
- Kripke, S. (1980) *Naming and Necessity* (Cambridge, MA: Harvard University Press).
- Levy, N. (2008) "Counterfactual Intervention and Agents' Capacities." *Journal of Philosophy* 105, 5: 223-39.
- Lewis, D. (1986) "Causation." In *Philosophical Papers II* (New York: Oxford University Press), pp. 159-213.
- McKenna, M. (2013) "Reasons-Responsiveness, Agents, and Mechanisms." In D. Shoemaker (ed.), *Oxford Studies in Agency and Responsibility*, vol. 1 (Oxford: Oxford University Press), pp. 151-83.
- Mele, A. (1997) "Introduction." In Mele (ed.), *The Philosophy of Action* (Oxford: Oxford University Press), pp. 1-26.
- Moya, C. (2018) "Reasons-Sensitivity, Causes, and Counterfactuals." *Teorema* 37, 1: 77-91.

- O'Connor, T. (2000) *Persons and Causes: The Metaphysics of Free Will* (Oxford: Oxford University Press).
- Pereboom, D. (2014) *Free Will, Agency, and Meaning in Life* (Oxford: Oxford University Press).
- Pereboom, D. (2018) "On Carolina Sartorio's *Causation and Free Will*." *Philosophical Studies* 175, 6: 1535-43.
- Rodriguez-Pereyra, G. (2015) "Grounding is not a Strict Order." *Journal of the American Philosophical Association* 1, 3: 517-34.
- Sartorio, C. (2009) "Omissions and Causalism." *Nous* 43, 3: 513-30.
- Sartorio, C. (2016) *Causation and Free Will* (Oxford: Oxford University Press).
- Sartorio, C. (2017a) "Frankfurt-Style Examples." In M. Griffin, N. Levy, and K. Time (eds.), *Routledge Companion to Free Will* (New York: Routledge), pp. 179-90.
- Sartorio, C. (2017b) "The Puzzle(s) of Frankfurt-Style Omission Cases." In D. Nelkin and S. Rickless (eds.), *The Ethics and Law of Omissions* (New York: Oxford University Press), pp. 133-47.
- Sartorio, C. (2018a) "Replies to Critics." *Teorema* 27, 1: 107-122.
- Sartorio C. (2018b) "Replies to Critics." *Philosophical Studies* 175, 6: 1545-56.
- Tognazzini, N. (2016) Review of *Causation and Free Will*, by C. Sartorio. *Journal of Philosophy* 113, 8: 417-22.
- Whittle, A. (2018) "Causation and the Grounds of Freedom." *Teorema* 37, 1: 61-76.